

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平6-259565

(43)公開日 平成6年(1994)9月16日

(51)Int.Cl.<sup>6</sup>

G 0 6 F 15/70

識別記号

4 6 5 A 8837-5L

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 2 F D (全 6 頁)

(21)出願番号 特願平5-75072

(22)出願日 平成5年(1993)3月10日

(71)出願人 000102728

エヌ・ティ・ティ・データ通信株式会社  
東京都江東区豊洲三丁目3番3号

(72)発明者 出本 浩

東京都江東区豊洲三丁目3番3号 エヌ・  
ティ・ティ・データ通信株式会社内

(72)発明者 木田 博巳

東京都江東区豊洲三丁目3番3号 エヌ・  
ティ・ティ・データ通信株式会社内

(74)代理人 弁理士 鈴木 正剛

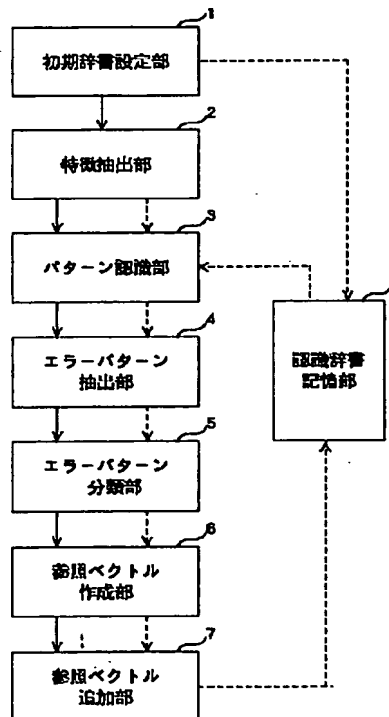
(54)【発明の名称】 パターン認識辞書作成システム

(57)【要約】

【目的】 認識率の高いパターン認識辞書を再構成するシステムを提供する。

【要約】

誤読された学習パターンをエラーパターン分類部5で正解側カテゴリと誤読先カテゴリのペアで分類してグループ化する。参照ベクトル作成部6では、これらグループ毎のパターンの平均ベクトルを求め、この平均ベクトルにおいて正解カテゴリの参照ベクトルに近いベクトル要素は正解カテゴリの参照ベクトル側の特徴値を、また誤読先カテゴリの参照ベクトルに近いベクトル要素は平均ベクトルの特徴値を持たせた参照ベクトルを新たに作成する。この参照ベクトルは、夫々の誤読された学習パターン特有の変形を表現している。そこでこの参照ベクトルを、参照ベクトル追加部7で既存のパターン認識辞書に追加する。



Best Available Copy

## 【特許請求の範囲】

【請求項1】 入力パターンの認識時に照合される参照ベクトルの集合からなる既存のパターン認識辞書を再構成するシステムであって、前記パターン認識辞書を初期辞書として設定する辞書設定手段と、前記入力パターンから特徴抽出を行う特徴抽出手段と、前記初期辞書を用いて前記入力パターンの特徴ベクトルのカテゴリを認識するパターン認識手段と、このパターン認識手段で認識された入力パターンから誤読されたものを抽出するパターン抽出手段と、抽出された誤読入力パターンを正解側カテゴリと誤読先カテゴリの組合わせ毎に分類してグループ化する分類手段と、各グループ毎に異なる値に基づく新たな参照ベクトルを作成する参照ベクトル作成手段と、この新たに作成された参照ベクトルを前記パターン認識辞書に追加する参照ベクトル追加手段と、を有することを特徴とするパターン認識辞書作成システム。

【請求項2】 前記参照ベクトル作成手段は、同一グループに含まれる入力パターン群の各特徴値の平均値列からなる平均ベクトルを求め、この平均ベクトルが正解側カテゴリ参照ベクトル及び誤読先カテゴリ参照ベクトルのいずれに近いかを前記特徴値毎に判定すると共に、誤読先カテゴリに近い特徴値については前記平均値、それ以外については正解側カテゴリ参照ベクトルの特徴値を夫々採用して参照ベクトルを作成するベクトル判定手段を含んでなることを特徴とする請求項1記載のパターン認識辞書作成システム。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、文字パターンや音声パターン等のパターン認識用の辞書作成システムに関し、特に、既存のパターン認識用の辞書（パターン認識辞書）を再構築して高精度化するために利用されるパターン認識辞書作成システムに関する。

## 【0002】

【従来の技術】 文字や図形あるいは音声を、例えば各種の制御や処理のための情報としてコンピュータに入力する場合、これら文字などを機械的に認識させる必要がある。このため、一般的には、文字等を例えば複数の特徴値の列からなるベクトル（以下「特徴ベクトル」という。）によって表現してパターン化し、これら文字パターン等の入力パターンに基づくパターン認識により、入力された文字等を判別する手法が採られている。

【0003】 上記のようなパターン認識においては、予め各カテゴリ（例えば文字パターン認識の場合には文字種類）の標準的なパターンを参照ベクトルとしてパターン認識辞書に用意しておく。そして、認識させたい入力パターンとパターン認識辞書における各カテゴリの参照ベクトルとの類似性を所定の評価関数を用いて計算し、この計算結果に基づいて、最も類似性の高いカテゴリを入力パターンの認識結果として出力する。

【0004】 また、文字や図形あるいは音声は、同じカテゴリでも種々の態様があるため、上記の参照ベクトルを作成する場合には、まずカテゴリごとにいくつかの異なるパターンの平均を計算し、これをパターン認識辞書に用意する。そして、文字認識等に携わる作業があるいは所謂エキスパートシステムによる学習等によって、このパターン認識辞書を再構成して高精度化し、パターン認識辞書による入力パターンの認識率を高めている。このようなパターン認識辞書を再構成するための技術として、従来、以下に挙げるような方法が提案されている。

【0005】 （1）第一の方法では、既存のパターン認識辞書により学習用の入力パターン（以下「学習パターン」という。）を認識させる。そして、入力パターンの誤読（誤認識）が最小になるように参照ベクトルの位置を最適化する処理を繰り返し行うことで、パターン認識辞書を再構成する。参照ベクトルの最適化の方法としては、例えば、LSIのレイアウト設計などで適用されているシミュレーテッドアニーリング法を採用したものが提案されている（松永、阿部、木田：“シミュレーテッドアニーリング法を用いた文字認識辞書の最適化”、信学技法、PRU90-39（1990））。

【0006】 （2）第二の方法は、既存のパターン認識辞書で学習パターンを認識させる点は上記方法と同じであるが、正しく認識できなかった入力パターンをそのまま対応するカテゴリの参照ベクトルとしてパターン認識辞書に新規追加することで、パターン認識辞書を再構成する。

【0007】 （3）第三の方法は、上記第二の方法に類似するが、パターン認識辞書のサイズをできる限り小さくするため、誤読された入力パターン1つに対して1つの参照ベクトルをパターン認識辞書に設けるのではなく、誤読された複数の入力パターンの平均を対応するカテゴリの参照ベクトルとしてパターン認識辞書に追加する（井出、若原：“文字認識におけるサブカテゴリ生成法の一検討”、1990年電子情報通信学会秋季全国大会論文集、D-348、p. 6-350（1990））。

## 【0008】

【発明が解決しようとする課題】 しかしながら、上記従来技術には、以下のような解決すべき課題があった。

【0009】 まず、第一の方法は、カテゴリ間のパターン分布の違いを考慮して認識精度が高くなるようにパターン認識辞書を最適化するという点では優れている。しかし、パターン分布が複雑でカテゴリ間でパターン分布が入り組み合っている場合、あるいはパターン分布が連続でなく分離（分散）している場合には、1つの参照ベクトルでは対処できない。このため、このようなパターンの場合には、パターン認識辞書の精度向上が図り難く、逆に認識精度の低下を招いてしまう。

【0010】また、第二の方法では、カテゴリごとに複数の参照ベクトルを持たせることができるため、第一の方法による問題点を解決することができる。しかし、この方法では、作られた参照ベクトルが学習時の入力パターンそのものに対応したものであるため、当該カテゴリのパターン認識用としての汎用性に乏しい。また、学習パターンを参照ベクトルとするため、その学習パターンが他のカテゴリにも類似している場合には、入力パターンを当該他のカテゴリのパターンと誤読する虞がある。

【0011】更に第三の方法では、複数のパターンの平均を参照ベクトルとすることで、パターン認識に際しての汎用性確保ができる。しかし、この方法では、学習時に誤読された入力パターンが少数の場合に、参照ベクトルにおける汎用性確保が難しく、誤読される入力パターンが多数無ければ効果が出ないという欠点がある。

【0012】本発明は、上記課題に鑑みてなされたもので、その目的とするところは、学習時において誤読される入力パターンが少数の場合でも汎用性が高い参照ベクトルが作成でき、認識率の高いパターン認識辞書を効率良く作成できる、パターン認識辞書作成システムを提供することにある。

【0013】

【課題を解決するための手段】上記目的を達成する本発明のパターン認識辞書作成システムは、入力パターンの認識時に照合される参照ベクトルの集合からなる既存のパターン認識辞書を再構成するシステムであって、前記パターン認識辞書を初期辞書として設定する辞書設定手段と、前記入力パターンから特徴抽出を行う特徴抽出手段と、前記初期辞書を用いて前記入力パターンの特徴ベクトルのカテゴリを認識するパターン認識手段と、このパターン認識手段で認識された入力パターンから誤読されたものを抽出するパターン抽出手段と、抽出された誤読入力パターンを正解側カテゴリと誤読先カテゴリの組合わせ毎に分類してグループ化する分類手段と、各グループ毎に異なる値に基づく新たな参照ベクトルを作成する参照ベクトル作成手段と、この新たに作成された参照ベクトルを前記パターン認識辞書に追加する参照ベクトル追加手段と、を有し、誤読パターン特有の変形を表現する参照ベクトルを作成して汎化能力の高いパターン認識辞書を得る構成とした。

【0014】上記構成において、参照ベクトル作成手段は、同一グループに含まれる入力パターン群の各特徴値の平均値列からなる平均ベクトルを求め、この平均ベクトルが正解側カテゴリ参照ベクトル及び誤読先カテゴリ参照ベクトルのいずれに近いかを特徴値毎に判定すると共に、誤読先カテゴリに近い特徴値については前記平均値、それ以外については正解側カテゴリ参照ベクトルの特徴値を夫々採用して参照ベクトルを作成するベクトル判定手段を含んでなる。

【0015】

【作用】上記構成である本発明のパターン認識辞書作成システムでは、まず、誤読された入力パターンを正解側カテゴリと誤読先カテゴリの組合わせでグループ化することで、類似した誤読傾向を持つもの同士を集める。そして、同じグループに含まれる入力パターンから得られた平均ベクトルにおいて、正解側カテゴリの参照ベクトルに近い特徴値は正解側カテゴリ参照ベクトルの値を、一方、誤読先カテゴリの参照ベクトルに近い特徴値は該平均ベクトルの平均値を夫々採用することで、誤読された入力パターンに共通した変形のみが異なって表現された汎用性の高い参照ベクトルが作成される。この新たに作成される参照ベクトルは、誤読される入力パターンが少数の場合でも有効であるため、この参照ベクトルをパターン認識辞書に追加することで、パターン認識辞書における認識率が向上する。

【0016】

【実施例】以下、本発明の実施例を図面を参照して説明する。

【0017】図1は、本発明の一実施例に係るパターン認識用辞書作成システムのブロック構成図である。図中、実線の矢印は、このシステムにおける処理の流れを示し、破線の矢印は、データの流れを示している。

【0018】図1を参照すると、本実施例のパターン認識辞書作成システムは、初期辞書設定部1、特徴抽出部2、パターン認識部3、エラーパターン抽出部4、エラーパターン分類部5、参照ベクトル作成部6、参照ベクトル追加部7、並びに認識辞書記憶部8から構成される。

【0019】初期辞書設定部1は、既存のパターン認識辞書を初期辞書として設定する（辞書設定手段）。このパターン認識辞書には、各カテゴリ毎の標準的なパターンが参照ベクトルとして夫々用意されている。特徴抽出部2は、入力された学習パターンの特徴抽出を行う（特徴抽出手段）。この抽出処理は、例えばパターン認識対象が文字の場合には、スキャナや文字切出し装置等の読取り手段によって学習対象となる文字ないし文字列を文字単位に2値化して入力し、これら入力された文字を複数のベクトル要素たる特徴値の配列からなる特徴ベクトルで表現して学習パターンとすることで行われる。

【0020】パターン認識部3では、初期辞書設定部1に設定された初期辞書を用いて学習パターンを認識する（パターン認識手段）。この認識処理は、初期辞書から取出した参照ベクトルと学習パターンの特徴ベクトルを、これらベクトルを構成する特徴値の配列等を照合して配列等の類似度の高いものを求めることにより行われ、これにより学習パターンの特定のカテゴリが認識される。エラーパターン抽出部4は、パターン認識部3において正しく認識されずに誤読された学習パターンを抽出する（パターン抽出手段）。このような抽出処理は、従来と同様に、例えば文字認識等に携わる作業者により

行われる。

【0021】エラーパターン分類部5は、上記抽出された誤読の学習パターンを、正しいカテゴリである正解側カテゴリと、誤読されたカテゴリである誤読先カテゴリとの組合わせ毎に分類してグループ化する（分類手段）。参照ベクトル作成部6は、これら分類されたグループ毎に、参照ベクトルを作成する（参照ベクトル作成手段）。この参照ベクトルについては後述する。この新たに作成された参照ベクトルは、参照ベクトル追加部7から認識辞書記憶部8に送出され、認識辞書記憶部8において初期辞書に追加される（参照ベクトル追加手段）。

【0022】次に、本実施例によるパターン認識辞書の再構成作業の処理手順を具体的に説明する。図2は上記処理手順を示すフローチャートであり、Sは各ステップを表す。図2を参照すると、まず、初期辞書設定部1で初期辞書を認識辞書記憶部8に記憶する（S101）。次いで、特徴抽出部2で学習パターンの特徴抽出を行い（S102）、パターン認識部3で初期辞書による学習パターンの認識を行う（S103）。エラーパターン抽出部4では、上記認識された学習パターンから誤読されたパターンを抽出し（S104）、次いでエラーパターン分類部5において、これら誤読された学習パターンが正解側カテゴリと誤読先カテゴリとのペアからなる複数のグループ $G_e$  ( $e=1, 2, \dots, E$ ) に分類する（S105）。この処理により、例えば、 $G_1$  は「亜」の学習パターンのうち「王」に誤読された学習パターンのグループ、 $G_2$  は「東」の学習パターンのうち「東」に誤読された学習パターンのグループとなる。

【0023】上記分類されたグループは、参照ベクトル作成部6においてグループ $G_e$  毎に平均ベクトル $m_{Ge}(i)$  ( $i=1, \dots, n$ ) が作成される（S107）。次いで参照ベクトル作成部6において、この作成された平均ベクトルが、正解側カテゴリ（例えばグループ $G_1$  であれば「亜」）の参照ベクトルと誤読先カテゴリ（同・「王」）の参照ベクトルとのいずれに近いかを、ベクトル要素、即ち特徴値毎に判定する（S108）。

【0024】この判定の結果、誤読先カテゴリに近い特徴値については上記作成された平均ベクトル $m_{Ge}(i)$  ( $i=1, \dots, n$ ) の平均値が（S109）、それ以外の正解カテゴリに近い特徴値については正解側カテゴリの参照ベクトル $m_c(i)$  ( $i=1, \dots, n$ ) の特徴値が（S110）、夫々採用される。この処理を全ての特徴値について行なうことで（S111）、グループ $G_e$  の新たな参照ベクトル $m'_{Ge}(i)$  ( $i=1, \dots, n$ ) を作成する。こうして作成した新たな参照ベクトル $m'_{Ge}(i)$  ( $i=1, \dots, n$ ) を、参照ベクトル追加部7により初期辞書に追加し（S112）、認識記憶部8に

記憶する。この工程を全てのグループ $G_1, G_2, \dots, G_E$  について行う（S106）。

【0025】このようにすれば、夫々の学習パターン特有の変形を表現する新たな参照ベクトルが逐次パターン認識辞書に追加され、パターン分布がカテゴリ間に入り組み合っている場合や、連続でなく分離している場合の認識精度の低下が抑制される。また、汎化能力も高まり、他カテゴリのパターンを引き込むことも無くなる。しかも、学習の際に誤読される入力パターンが少数でも汎用性が高い参照ベクトルが作成できることから、認識率の高いパターン認識辞書を効率よく作成できる。なお、本発明はこの実施例に拘束されるものではなく、その要旨を逸脱しない範囲で図1の構成あるいは図2の処理手順の変更が可能である。

#### 【0026】

【発明の効果】以上詳細に説明したように、本発明のパターン認識辞書作成システムでは、正解側と誤読先のペアにて誤読の学習パターンをグループ化するとともに、各グループに含まれるパターン群について、特徴値の平均値の列からなる平均ベクトルを求め、該平均ベクトルが正解カテゴリ参照ベクトル及び誤読先カテゴリ参照ベクトルのいずれかに近いかを特徴値毎に判定し、誤読先に近いベクトル要素については該平均値、それ以外については正解側カテゴリ参照ベクトルの特徴値を採用して新たな参照ベクトルを形成し、この新たな参照ベクトルを既存のパターン認識辞書に追加するようにしたので、夫々のエラー学習パターン特有の変形を表現する参照ベクトルを形成することが可能となり、汎化能力の高い認識辞書が得られる。

【0027】また、学習の際に誤読される入力パターンが少数でも汎用性が高い参照ベクトルが作成できることから、認識率の高いパターン認識辞書を効率よく作成でき、既存のパターン認識辞書の高精度化のためのパターン認識辞書の再構成が効率良く行える。

#### 【図面の簡単な説明】

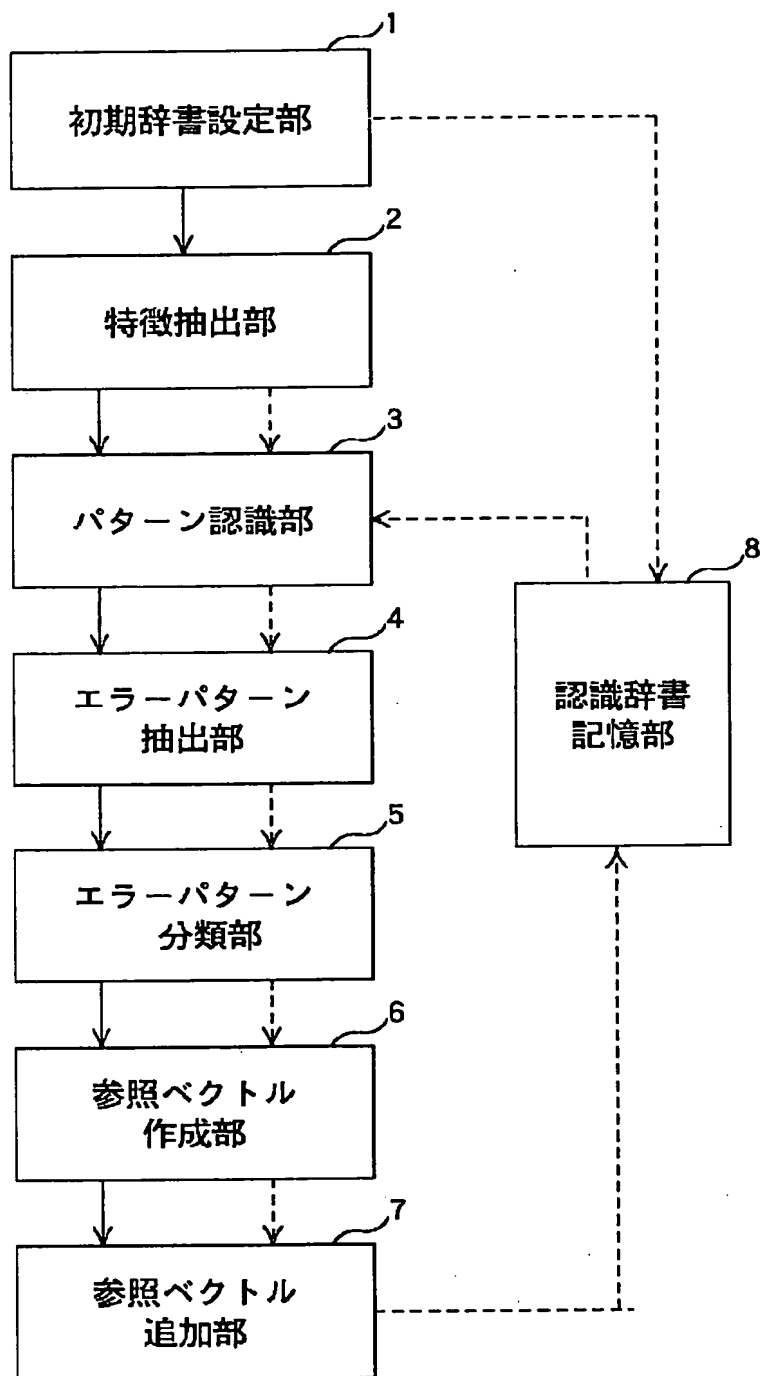
【図1】本発明の一実施例に係るパターン認識辞書作成システムの要部構成図。

【図2】この実施例におけるパターン認識辞書作成の処理手順を示すフローチャート。

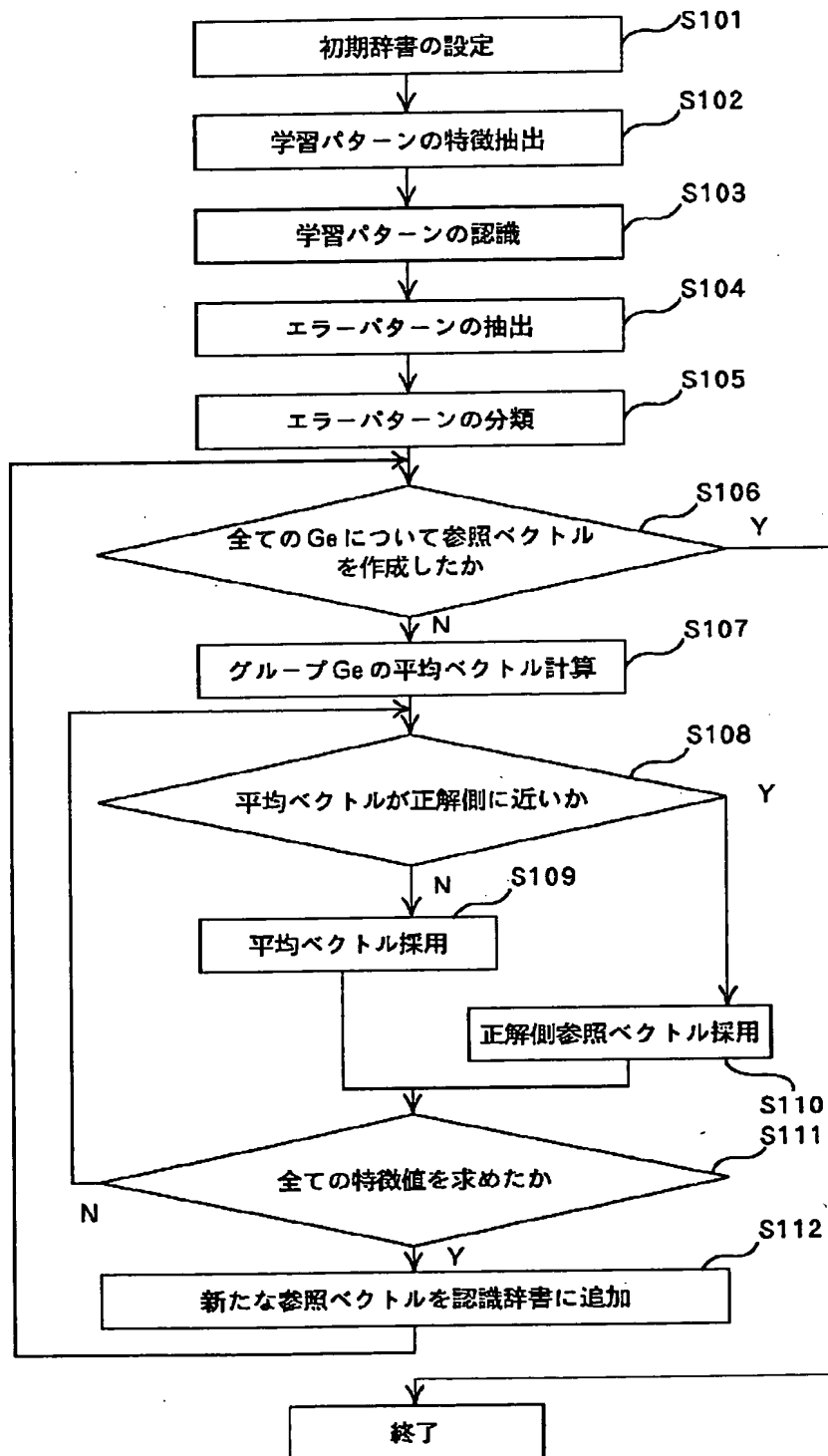
#### 【符号の説明】

- 1 初期辞書設定部（辞書設定手段）
- 2 特徴抽出部（特徴抽出手段）
- 3 パターン認識部（パターン認識手段）
- 4 エラーパターン抽出部（パターン抽出手段）
- 5 エラーパターン分類部（分類手段）
- 6 参照ベクトル作成部（参照ベクトル作成手段）
- 7 参照ベクトル追加部（参照ベクトル追加手段）
- 8 認識辞書記憶部

【図1】



【図2】



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**